Overestimation of Postpartum Depression Prevalence Based on a 5-item Version of the EPDS: Systematic Review and Individual Participant Data Meta-Analysis

Brett D. Thombs, PhD[1-7,*], Brooke Levis, PhD[1,2,8], Anita Lyubenova[1], Dipika Neupane, BPH[1,2], Zelalem Negeri, PhD[1,2], Yin Wu, PhD[1,2,4], Ying Sun, MPH[1], Chen He, MScPH[1,2], Ankur Krishnan, MSc[1], Simone N. Vigod, MD[9], Parash Mani Bhandari, BPH[1,2], Mahrukh Imran, MScPH[1], Danielle B. Rice, MSc[1,3], Marleine Azar, MSc[1,2], Matthew J. Chiovitti, MISt[1], Nazanin Saadat, MSc[1], Kira E. Riehm, MSc[1,10], Jill T. Boruff, MLIS[11], Pim Cuijpers, PhD[12], Simon Gilbody, PhD[13], John P. A. Ioannidis, MD[14], Lorie A. Kloda, PhD[15], Scott B. Patten, PhD[16-18], Ian Shrier, MD[1,2,19], Roy C. Ziegelstein, MD[20], Liane Comeau, PhD[21], Nicholas D. Mitchell, MD[22,23], Marcello Tonelli, MD[24], Jacqueline Barnes, PhD[25], Cheryl Tatano Beck, DNSc[26], Carola Bindt, MD[27], Barbara Figueiredo, PhD[28], Nadine Helle, PhD[27], Louise M. Howard, PhD[29,30], Jane Kohlhoff, PhD[31-33], Zoltán Kozinszky, MD[34], Angeliki A. Leonardou, MD[35], Sandra Nakić Radoš, PhD[36], Chantal Quispel, PhD[37], Tamsen J. Rochat, PhD[38,39], Alan Stein, FRCPsych[40,41], Robert C. Stewart, PhD[42,43], Meri Tadinac, PhD[44], S. Darius Tandon, PhD[45], Iva Tendais, PhD[28], Annamária Töreki, PhD[46], Thach D. Tran, PhD[47], Kylee Trevillion, PhD[29], Katherine Turner, PsyD[48], Johann M. Vega-Dienstmaier, MD[49], Andrea Benedetti, PhD[2,5,50,*]

* Co-senior authors

Affiliations:

[1]Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada; [2]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada; [3]Department of Psychology, McGill University,

Montréal, Québec, Canada; [4]Department of Psychiatry, McGill University, Montréal, Québec, Canada; [5]Department of Medicine, McGill University, Montréal, Québec, Canada; [6]Department of Educational and Counselling Psychology, McGill University, Montréal, Québec, Canada; [7]Biomedical Ethics Unit, McGill University, Montréal, Québec, Canada; [8]Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Staffordshire, UK; [9]Women's College Hospital and Research Institute, University of Toronto, Toronto, Ontario, Canada; [10]Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA; [11]Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montréal, Québec, Canada; [12]Department of Clinical, Neuro and Developmental Psychology, EMGO Institute, Vrije Universiteit Amsterdam, the Netherlands; [13]Hull York Medical School and the Department of Health Sciences, University of York, Heslington, York, UK; [14]Department of Medicine, Department of Epidemiology and Population Health, Department of Biomedical Data Science, Department of Statistics, Stanford University, Stanford, California, USA; [15]Library, Concordia University, Montréal, Québec, Canada; [16]Departments of Community Health Sciences and Psychiatry, University of Calgary, Calgary, Canada; [17]Mathison Centre for Mental Health Research & Education, University of Calgary, Calgary, Canada; [18]Cuthbertson & Fischer Chair in Pediatric Mental Health, University of Calgary, Calgary, Canada; [19]Department of Family Medicine, McGill University, Montréal, Québec, Canada; [20]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [21]International Union for Health Promotion and Health Education, École de santé publique de l'Université de Montréal, Montréal, Québec, Canada; [22]Department of Psychiatry, University of Alberta, Edmonton, Alberta, Canada; [23]Alberta Health Services, Edmonton, Alberta, Canada; [24]Department of Medicine, University of Calgary, Calgary, Alberta,

Canada; [25]Department of Psychological Sciences, Birkbeck, University of London, UK; [26]University of Connecticut School of Nursing, Mansfield, Connecticut, USA; [27]Department of Child and Adolescent Psychiatry, University Medical Center Hamburg-Eppendorf, Germany; [28]School of Psychology, University of Minho, Portugal; [29]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; [30]South London and Maudsley NHS Foundation Trust, London, UK; [31]School of Psychiatry, University of New South Wales, Kensington, Australia; [32]Ingham Institute, Liverpool, Australia; [33]Karitane, Carramar, Australia; [34]Department of Obstetrics and Gynaecology, Blekinge Hospital, Karlskrona, Sweden; [35]First Department of Psychiatry, Women's Mental Health Clinic, Athens University Medical School, Athens, Greece; [36]Department of Psychology, Catholic University of Croatia, Croatia; [37]Department of Obstetrics and Gynaecology, Albert Schweitzer Ziekenhuis, Dordrecht, the Netherlands; [38]MRC/Developmental Pathways to Health Research Unit, Faculty of Health Sciences, University of Witwatersrand, South Africa; [39]Human and Social Development Program, Human Sciences Research Council, South Africa; [40]Department of Psychiatry, University of Oxford, Oxford, UK; [41]MRC/Wits Rural Public Health and Health Transitions Research Unit, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; [42]Division of Psychiatry, University of Edinburgh, Edinburgh, Scotland; [43]Malawi Epidemiology and Intervention Research Unit (MEIRU), Lilongwe, Malawi; [44]Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia; [45]Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; [46]Department of Emergency, University of Szeged, Hungary; [47]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia; [48]Epilepsy Cter-Child Neuropsychiatry Unit, ASST Santi Paolo Carlo, San Paolo Hospital, Milan, Italy;

[49]Facultad de Medicina Alberto Hurtado, Universidad Peruana Cayetano Heredia. Lima, Perú; [50]Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, Québec, Canada.

Corresponding authors:

Brett D. Thombs, PhD; Jewish General Hospital; 4333 Cote Ste Catherine Road; Montreal, Quebec H3T 1E4; Tel (514) 340-8222 ext. 25112; E-mail: brett.thombs@mcgill.ca

Andrea Benedetti, PhD; Centre for Outcomes Research & Evaluation, Research Institute of the McGill University Health Centre, 5252 Boulevard de Maisonneuve, Montréal, QC, H4A 3S5, Canada; Tel (514) 934-1934 ext. 32161; E-mail: andrea.benedetti@mcgill.ca

Word count: 2,732

ABSTRACT

Objective: The Maternal Mental Health in Canada, 2018/2019 survey reported that 18% of 7,085 mothers who recently gave birth reported "feelings consistent with postpartum depression" based on scores ≥7 on a 5-item version of the Edinburgh Postpartum Depression Scale (EPDS-5). The EPDS-5 was designed as a screening questionnaire, not to classify disorders or estimate prevalence; the extent to which EPDS-5 results reflect depression prevalence is unknown. We investigated EPDS-5 ≥7 performance relative to major depression prevalence based on a validated diagnostic interview, the Structured Clinical Interview for DSM (SCID).

Methods: We searched Medline, Medline In-Process & Other Non-Indexed Citations, PsycINFO, and the Web of Science Core Collection through June 2016 for studies with datasets with item response data to calculate EPDS-5 scores and that used the SCID to ascertain depression status. We conducted an individual participant data meta-analysis to estimate pooled percentage of EPDS-5 ≥7, pooled SCID major depression prevalence, and the pooled difference in prevalence.

Results: 3,958 participants from 19 primary studies were included. Pooled prevalence of SCID major depression was 9.2% (95% confidence interval [CI] 6.0% to 13.7%), pooled percentage of participants with EPDS-5 ≥7 was 16.2% (95% CI 10.7% to 23.8%), and pooled difference was 8.0% (95% CI 2.9% to 13.2%). In the 19 included studies, mean and median ratios of EPDS-5 to SCID prevalence were 2.1 and 1.4 times.

Conclusions: Prevalence estimated based on EPDS-5 ≥7 appears to be substantially higher than the prevalence of major depression. Validated diagnostic interviews should be used to establish prevalence.

Keywords: Epidemiology, Evidence-based Medicine, Obstetrics and Gynecology, Psychiatry, Statistics and Research Methods

Depression during pregnancy and the postpartum period is associated with negative implications for maternal health, child health, and families.[1-3] Accurate estimation of depression prevalence in this population is important for understanding disease burden, making informed decisions regarding health care resources, and investigating etiology and challenges associated with the condition. Systematic reviews have reported postpartum depression prevalence as approximately 7% based on Diagnostic and Statistical Manual (DSM) criteria.[4,5] A study of over 14,000 women in the United States found that 8% of women in pregnancy and 9% of women within 12 months postpartum met DSM-IV criteria for depression based on a diagnostic interview, compared to 8% among same-aged women.[6]

The Maternal Mental Health in Canada, 2018/2019 survey reported that 18% of 7,085 mothers who gave birth between 5 and 13 months prior reported "feelings consistent with post-partum depression"[7] based on scoring ≥7on a 5-item version of the Edinburgh Postpartum Depression Scale (EPDS-5).[8] Self-report questionnaires, including the EPDS-5, include some symptoms used to diagnose depression, but they do not include all relevant symptoms, consideration of functional impairment, or information needed for differential diagnosis.[9-11]

Cutoff thresholds on screening tools are typically set to cast a wide net and identify people who may benefit from further evaluation, but not to determine whether diagnostic criteria are met or estimate prevalence.[9-11] Ascertainment of case status and prevalence estimation require the use of a validated diagnostic interview, such as the Structured Interview for DSM (SCID).[12] The 10-item EPDS is commonly researched. Less is known about the performance of the EPDS-5, which has been evaluated only in a single study of 56 women (9 depression cases). Knowledge about how it performs in a larger sample would greatly assist interpretation of Maternal Mental

Health in Canada results and inform recommendations about its use for describing disease burden.

The present study used data from an individual participant data meta-analysis (IPDMA) on EPDS depression screening tool accuracy to compare the proportion of women in pregnancy or postpartum with scores ≥7 on the EPDS-5 to prevalence of major depression based on the SCID.

METHODS

This study was conducted with data accrued for an IPDMA on EPDS screening accuracy. The original IPDMA was registered (PROSPERO; CRD42015024785), and a protocol was published.[13] The present study was not included in the main EPDS IPDMA protocol. It was conducted using methods from a similar study of prevalence based on the full EPDS with the protocol uploaded to the Open Science Framework prior to initiating analyses (https://osf.io/7gy6p/).

Identification of Eligible Studies

Datasets from articles in any language were eligible for the main IPDMA if (1) they included EPDS scores for women during pregnancy or within 12 months postpartum; (2) they included current Major Depressive Episode (MDE) or Major Depressive Disorder (MDD) classifications based on DSM[14-16] or International Classification of Diseases[17] criteria based on a validated semi-structured or fully structured interview; (3) the EPDS and interview were done within two weeks of each other; (4) participants were ≥18 years old and not recruited from school settings, since the database was originally accrued to assess screening accuracy among adults, and school-based screening may have different characteristics; and (5) participants were not recruited from psychiatric settings or because they were pre-identified as possibly having

depression. Datasets where not all participants were eligible were included if individual eligible participants could be identified.

In the present study we included only data from primary studies that based major depression diagnoses on the SCID.[12] It is intended for administration by a trained diagnostician, requires clinical judgment, and allows probes to be made to clarify responses. We only included studies that used the SCID because semi-structured interviews replicate diagnostic standards more closely than other types of interviews, and the SCID is by far the most commonly used semi-structured diagnostic interview for depression research.[18-20] Three previous analyses that used large IPDMA databases found that, compared to semi-structured interviews, fully structured interviews, designed for administration by lay interviewers, identified more participants with low-level depressive symptoms but fewer participants with high-level symptoms as depressed.[18-20] One brief fully structured interview, the Mini International Neuropsychiatric Interview, identified far more participants as being depressed across the symptom spectrum.[18-20] Additionally, we excluded datasets that provided only total EPDS scores without item scores. This is because item scores were needed to calculate EPDS-5 scores.

Data Sources, Search Strategy, and Study Selection

We searched Medline, Medline In-Process & Other Non-Indexed Citations and PsycINFO via OvidSP, and the Web of Science Core Collection via ISI Web of Knowledge from inception to June 10, 2016. The search was designed by an experienced medical librarian and peer-reviewed (Appendix 1).[21] We reviewed reference lists from published reviews and queried collaborators to attempt to identify non-published studies. Search results were uploaded into RefWorks (RefWorks-COS, Bethesda, MD, USA), and, after duplicate removal, into DistillerSR (Evidence Partners, Ottawa, Canada) for managing the review process and data extraction.

Two investigators independently reviewed titles and abstracts, and if either deemed a study potentially eligible, full-text review was done by two investigators, independently. Any disagreements were resolved by consensus, with a third investigator consulted if necessary.

Data Contribution and Synthesis

Authors of studies with eligible datasets were contacted and invited to contribute de-identified primary datasets. We emailed corresponding authors of eligible primary studies at least three times, with at least two weeks between each email. If there was not a response, we attempted phone contact and emailed co-authors.

For each contributed dataset, we attempted to verify that we could replicate published participant characteristics and screening accuracy results, and we resolved any discrepancies, consulting with the study investigators. The number of participants and cases from a primary study in the IPDMA dataset sometimes differed from numbers in published primary study reports for several reasons. First, for some primary studies, not all participants met inclusion criteria for our IPDMA. This occurred, for instance, if the period between administration of the EPDS and diagnostic interview was longer than two weeks for some participants. Second, some primary studies reported accuracy results for depression diagnoses broader than major depression, such as “any depressive disorder”, but our reference standard was major depression, which would have resulted in a different number of cases than published. Third, in some cases, when we compared published results with results from contributed datasets, there were discrepancies, and we used the corrected results.

For primary datasets that used sampling procedures that required weighting, we used the weights provided. This occurred, for instance, in studies where all participants with positive screens and a random subset of participants with negative screens received a diagnostic

interview. For studies where sampling should have been done, but weights were not available, we used inverse selection probabilities.

Statistical Analyses

For each primary study, we calculated the prevalence of major depression based on the SCID, the percentage who scored ≥7 on the EPDS-5, the difference in prevalence between the two methods (EPDS-5 ≥ 7 prevalence – SCID major depression prevalence), and the corresponding ratio. Then, across studies, we pooled (1) percentage with EPDS-5 ≥7, (2) prevalence of SCID major depression, and (3) the differences in prevalence from each study. We also determined the mean and median ratio for EPDS-5 ≥7 versus SCID major depression prevalence.

All meta-analyses were conducted in R (R version 3.4.1; R Studio version 1.0.143) using the lme4 package. Given the clustered nature of the data, mixed-effects models were used. To estimate pooled prevalence values, generalized linear mixed-effects models with a logit link function were fit using the glmer function. The logit link accounts for the binary nature of the outcome (EPDS-5 ≥7 vs <7; presence vs. absence of SCID major depression). To estimate the pooled difference value (fit continuously, given that differences could be positive or negative), a linear mixed-effects model was fit using the lmer function. In all analyses, to account for correlation between subjects within the same primary study (i.e., the clustering), random intercepts were fit for each primary study. To quantify heterogeneity, for each analysis, we (1) calculated τ^2, which is the estimate of between-study variance, (2) calculated the I^2 statistic, which quantifies the proportion of total variability due to between-study heterogeneity, and (3) estimated the 95% prediction interval for the difference in prevalence, which illustrates the range

of difference values that would be expected if a new study were to compare proportion with EPDS-5 ≥7 to prevalence based on SCID.

In post-hoc analyses, we investigated whether differences in prevalence (EPDS-5 ≥7 prevalence – SCID major depression prevalence) were associated with study and participant characteristics. To do this, we fit additional linear mixed-effects models for pooled prevalence difference, including age, pregnant versus postpartum status, country human development index ("very high", "high", or "low-medium", based on the United Nation's Human Development Index for the year of publication), and study sample size as fixed-effect covariates.

Ethical Approval

Since this study involved analysis of previously collected de-identified data and because included studies were required to have obtained ethics approval and informed consent, the Research Ethics Committee of the Jewish General Hospital determined that ethics approval was not required.

RESULTS

Search Results and Inclusion of Primary Study Datasets

There were 3,417 unique citations identified, of which 3,097 were excluded after review of titles and abstracts and 212 after full-text review. The 108 remaining articles comprised data from 73 unique samples, of which 49 provided data for the main IPDMA; in addition, we were provided data from one unpublished study, which was subsequently published. For the present study, of the 50 study datasets in the main IPDMA, 21 were excluded because they used a diagnostic interview other than the SCID (19 fully structured interviews, 2 other semi-structured interviews), and 10 were excluded because item-level data to calculate EPDS-5 scores were not available. Thus, datasets from 19 studies were included with 3,958 participants (572 cases of

major depression; prevalence 14%). Figure 1 shows the search and dataset inclusion processes, and Table 1 shows the characteristics of each included study.[22-40]

Depression Prevalence Based on the SCID Versus EPDS-5 ≥7

The pooled prevalence of SCID major depression was 9.2% (95% confidence interval [CI]: 6.0% to 13.7%; τ^2: 0.901; I^2: 94.4%). The pooled percentage of participants who scored ≥7 on the EPDS-5 was 16.2% (95% CI: 10.7% to 23.8%; τ^2: 1.044; I^2: 94.6%). The pooled difference from each study was 8.0% (95% CI: 2.9% to 13.2%; τ^2: 0.010; I^2: 93.7%; 95% prediction interval: -13.8% to 29.9%). In the 19 included primary studies, the mean and median ratios of proportion of EPDS-5 ≥7 versus SCID prevalence were 2.1 and 1.4, respectively. See Table 1.

In post-hoc analyses, no study or participant characteristics were significantly associated with differences in prevalence, with the exception of age, for which a one-year increase in age was associated with a 0.4% (95% CI: 0.2% to 0.7%) decrease in "EPDS-5 ≥4 minus SCID" prevalence.

DISCUSSION

The Maternal Mental Health in Canada, 2018/2019 survey was conducted by Statistics Canada in collaboration with the Public Health Agency of Canada and Health Canada in order to address a pressing need for data on maternal mental health problems, including depression.[7] One previous study had suggested that the EPDS-5 with a cutoff of ≥7 could be used as a screening tool for depression, but it was based on only 9 cases and did not attempt to calibrate the tool to estimate prevalence. Results from the present analysis suggest that using a score of ≥7 on the EPDS-5 overestimates true prevalence by an absolute value of about 8% or approximately 1.4 to 2.0 times, depending on whether a mean or median ratio of EPDS-5 to SCID prevalence is used.

Despite the heterogeneity across studies in our IPDMA, it is safe to conclude that depression prevalence would be substantially overestimated by an EPDS-5 cutoff of ≥7, although it is less easy to determine the amount of overestimation in any given study. This finding is similar to other studies that have found that estimates of prevalence derived from cutoff scores on screening scales used clinically to detect patients with possible depression vastly over-estimate prevalence by diagnostic interview.[10,11]

The implication of using terminology such as "feelings consistent with post-partum depression," as used in Maternal Mental Health in Canada, 2018/2019 survey is also important. Diagnostic or classification thresholds are set to identify individuals with a condition or level of impairment that warrants medical attention. Although women who score ≥7 on the EPDS-5 have symptoms that are on average more consistent with depression than those below that threshold, this does not necessarily mean that they have a diagnosis of depression or require treatment, making it very difficult to use the information generated, other than perhaps to compare symptom burden across other populations or samples using similar thresholds on the same scale.

The overestimation of prevalence may also have implications beyond assessing depression prevalence itself. For example, the Maternal Mental Health in Canada survey reported that 12% of women who were classified as depressed with EPDS ≥7 had experienced thoughts of harming themselves "sometimes" or "often" since the birth of their child. Since many more women were classified as depressed than would have met diagnostic criteria based on a validated interview, it is possible that the true proportion of women with major depression with thoughts of self-harm could be substantially greater than what was estimated. Misclassification not only affects our understanding of the frequency of a condition but also how we understand the experiences and challenges of those with the condition.

There are many examples of national surveys that have used validated diagnostic interviews to estimate depression prevalence. In Canada, the Canadian Community Health Survey – Mental Health used a version of the World Health Organization's fully structured Composite International Diagnostic Interview (CIDI) to evaluate the prevalence of major depressive disorder with a sample of over 25,000 participants.[41] In the United States, the Epidemiologic Catchment Area Study used another fully structured interview, the Diagnostic Interview Schedule (DIS),[42] and the National Comorbidity Survey used the CIDI.[43] Large cross-national studies have similarly used the DIS[44] and the CIDI.[45] The use of validated diagnostic interviews requires substantial resources. Using alternative methods, such as the EPDS-5, which over-identify depression cases, however, makes it difficult to understand where needs are greatest, identify factors associated with onset of mental health problems, and find effective solutions. When resources are not available to properly identify cases, alternative research questions can be considered.

STRENGTHS AND LIMITATIONS

An important strength of the present study is that it included data from 19 primary studies with almost 4,000 participants and almost 600 cases of major depression based on the SCID, a rigorous semi-structured diagnostic interview designed to classify psychiatric disorders, including major depression. We were able to directly compare the proportion of women with EPDS-5 ≥7 and prevalence of major depression based on the SCID. A limitation was that included studies came from many different countries and reported different prevalence of major depression, although the pooled percentage of participants with EPDS-5 ≥7 (16%) was similar to that of the Maternal Mental Health in Canada, 2018/2019 survey (18%). Another was that the search included studies only through June 2016. There was also considerable heterogeneity

across studies in the difference between prevalence estimated with EPDS-5 ≥7 versus the SCID. Although age was statistically significantly associated with the difference between EPDS-5 ≥7 prevalence and SCID major depression prevalence, a one-year difference was associated with only a 0.4% difference; given the general similarity in ages of pregnant and postpartum women, this would not explain the large differences we found.

Despite these limitations, there was robust evidence that the EPDS-5 ≥7 generally overestimates depression prevalence, and that the magnitude of the overestimation appears to be clinically important.

CONCLUSIONS

In summary, we found that using EPDS-5 ≥7 to estimate depression overestimates the true prevalence of depression substantially. As such, while the 18% reported in the Maternal Mental Health in Canada, 2018/2019 survey reflects a certain burden of depressive symptomatology, policy-makers may not be able to use it as a benchmark for planning levels of specific services because many of those scoring 7 or above on a scale such as the EPDS-5 would not be diagnosed with major depressive disorder in a clinical interview. Postpartum depression is an important and burdensome condition, and as such, future surveys should use validated diagnostic interviews designed for diagnostic calibration to understand prevalence and provide more accurate data to use as a benchmark for policymakers to be able to act on need for service to improve outcomes for affected mothers and children.

Contributors: BDT, BL, AL, JBoruff, PC, SG, JPAI, LAK, SBP, IS, RCZ, LC, NDM, MTonelli, SNV, and AB were responsible for the study conception and design. JBoruff and LAK designed and conducted database searches to identify eligible studies. JBarnes, CTB, CB, FPF, BF, NH, LMH, JK, ZK, AAL, SNR, CQ, TJR, AS, RCS, MTadinac, SDT, IT, AT, TDT, KTrevillion, KTurner, JMVD contributed primary datasets that were included in this study. BDT, BL, AL, DN, ZN, YW, YS, CH, DBR, AK, PMB, MA, MJC, NS, KER, and MI contributed to data extraction and coding for the meta-analysis. BDT, BL, and AB contributed to the data analysis and interpretation. BDT, BL, AL, and AB contributed to drafting the manuscript. All authors provided a critical review and approved the final manuscript. BDT and AB are the guarantors; they had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analyses.

Declaration of Competing Interests: All authors have completed the ICJME uniform disclosure form and declare: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years with the following exceptions: Dr. Tonelli declares that he has received a grant from Merck Canada, outside the submitted work. Dr. Vigod declares that she receives royalties from UpToDate, outside the submitted work. Dr. Beck declares that she receives royalties for her Postpartum Depression Screening Scale published by Western Psychological Services. Dr. Howard declares that she has received personal fees from NICE Scientific Advice, outside the submitted work. All authors declare no other relationships or activities that could appear to have influenced the submitted work. No funder had any role in the design and conduct

of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Funding: This study was funded by the Canadian Institutes of Health Research (CIHR, KRS-140994). Drs. Thombs and Benedetti were supported by Fonds de recherche du Québec - Santé (FRQS) researcher salary awards. Dr. Levis was supported by a FRQS Postdoctoral Training Fellowship. Ms. Lyubenova was supported by the Mitacs Globalink Research Internship Program. Ms. Neupane was supported by G.R. Caverhill Fellowship from the Faculty of Medicine, McGill University. Dr. Wu was supported by a FRQS Postdoctoral Training Fellowship. Ms. Rice was supported by a Vanier Canada Graduate Scholarship. Mr. Bhandari was supported by a studentship from the Research Institute of the McGill University Health Centre. Ms. Azar was supported by a FRQS Masters Training Award. The primary study by Barnes et al. was supported by a grant from the Health Foundation (1665/608). The primary study by Beck et al. was supported by the Patrick and Catherine Weldon Donaghue Medical Research Foundation and the University of Connecticut Research Foundation. The primary study by de Figueiredo et al. was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo. The primary study by Tendais et al. was supported under the project POCI/SAU-ESP/56397/2004 by the Operational Program Science and Innovation 2010 (POCI 2010) of the Community Support Board III and by the European Community Fund FEDER. The primary study by Helle et al. was supported by the Werner Otto Foundation, the Kroschke Foundation, and the Feindt Foundation. The primary study by Howard et al. was supported by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (Grant Reference Numbers RP-PG-1210-12002 and RP-DG-1108-10012) and by the

South London Clinical Research Network. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The primary study by Phillips et al. was supported by a scholarship from the National Health and Medical and Research Council (NHMRC). The primary study by Nakić Radoš et al. was supported by the Croatian Ministry of Science, Education, and Sports (134-0000000-2421). The primary study by Quispel et al. was supported by Stichting Achmea Gezondheid (grant number z-282). The primary study by Rochat et al. was supported by grants from the University of Oxford (HQ5035), the Tuixen Foundation (9940), the Wellcome Trust (082384/Z/07/Z and 071571), and the American Psychological Association. Dr. Rochat receives salary support from a Wellcome Trust Intermediate Fellowship (211374/Z/18/Z). The primary study by Prenoveau et al. was supported by The Wellcome Trust (grant number 071571). The primary study by Stewart et al. was supported by Professor Francis Creed's Journal of Psychosomatic Research Editorship fund (BA00457) administered through University of Manchester. The primary study by Tandon et al. was funded by the Thomas Wilson Sanitarium. The primary study by Tran et al. was supported by the Myer Foundation who funded the study under its Beyond Australia scheme. Dr. Tran was supported by an early career fellowship from the Australian National Health and Medical Research Council. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The primary study by Vega-Dienstmaier et al. was supported by Tejada Family Foundation, Inc, and Peruvian-American Endowment, Inc. No other authors reported funding for primary studies or for their work on the present study.

REFERENCES

1. Stewart DE. Perinatal depression. *Gen Hosp Psychiatry*. 2006;28:1-2.
2. Howard LM, Molyneaux E, Dennis C-L, et al. Non-psychotic mental disorders in the perinatal period. *Lancet*. 2014;384:1775-1788.
3. Stewart DE, Vigod S. Postpartum depression. *N Eng J Med*. 2016;375:2177-2186.
4. O'Hara MW, Swain AM. Rates and risk of postpartum depression-a meta-analysis. *Int Rev Psychiatr*. 1996;8(1):37-54.
5. Gavin NI, Gaynes BN, Lohr KN, et al. Perinatal depression: a systematic review of prevalence and incidence. *Obstet Gynecol*. 2005;106(5):1071-1083.
6. Vesga-López O, Blanco C, Keyes K, et al. Psychiatric disorders in pregnant and postpartum women in the United States. *Arch Gen Psychiatry*. 2008;65(7):805-15.
7. Maternal Mental Health in Canada, 2018/2019. Statistics Canada. 2019. Available from: https://www150.statcan.gc.ca/n1/daily-quotidien/190624/dq190624b-eng.htm. Accessed December 6, 2019.
8. Eberhard-Gran M, Eskild A, Samuelsen SO, Tambs K. A short matrix-version of the Edinburgh Depression Scale. *Acta Psychiatr Scand*. 2007;116(3):195-200.
9. Thombs BD, Kwakkenbos L, Levis AW, Benedetti A. Addressing overestimation of the prevalence of depression based on self-report screening questionnaires. *CMAJ*. 2018;190(2):E44-E49.
10. Levis B, Yan XW, He C, et al. Comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: a meta-research review. *BMC Med*. 2019;17(1):65.

11. Levis B, Benedetti A, Ioannidis JPA, et al. Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis. *J Clin Epidemiol*. 2020. doi: 10.1016/j.jclinepi.2020.02.002.
12. First MB, Gibbon M. The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II). In: Hilsenroth MJ, Segal DL, eds. *Comprehensive handbook of psychological assessment: Vol. 2. Personality assessment*. Hoboken, NJ: John Wiley & Sons Inc; 2004:134-143.
13. Thombs BD, Benedetti A, Kloda LA, et al. Diagnostic accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for detecting major depression in pregnant and postnatal women: protocol for a systematic review and individual patient data meta-analyses. *BMJ Open*. 2015;5(10):e009742.
14. Diagnostic and statistical manual of mental disorders: DSM-III 3rd ed, revised. Washington, DC: American Psychiatric Association 1987.
15. Diagnostic and statistical manual of mental disorders: DSM-IV 4th ed. Washington, DC: American Psychiatric Association 1994.
16. Diagnostic and statistical manual of mental disorders: DSM-IV 4th ed, text revised. Washington, DC: American Psychiatric Association 2000.
17. The ICD-10 Classifications of Mental and Behavioural Disorder: Clinical Descriptions and Diagnostic Guidelines Geneva: World Health Organization 1992.
18. Levis B, Benedetti A, Riehm KE, et al. Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. *Br J Psychiatry*. 2018;212(6):377-385.

19. Levis B, McMillan D, Sun Y, et al. Comparison of major depression diagnostic classification probability using the SCID, CIDI, and MINI diagnostic interviews among women in pregnancy or postpartum: An individual participant data meta-analysis. *Int J Methods Psychiatr Res*. 2019;28(4):e1803.

20. Wu Y, Levis B, Sun Y, et al. Probability of major depression diagnostic classification based on the SCID, CIDI and MINI diagnostic interviews controlling for Hospital Anxiety and Depression Scale – Depression subscale scores: an individual participant data meta-analysis of 73 primary studies. *J Psychosom Res*. 2020;129:109892.

21. PRESS – Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E). Ottawa: CADTH; 2016 Jan.

22. Barnes JS, Senior R, MacPherson K. The utility of volunteer home-visiting support to prevent maternal depression in the first year of life. *Child Care Health Dev*. 2009;35(6):807-816.

23. Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. *Nurs Res*. 2001;50(4):242-250.

24. de Figueiredo FP, Parada AP, Cardoso VC, et al. Postpartum depression screening by telephone: a good alternative for public health and research. *Arch Womens Ment Health*. 2015;18(3):547-553.

25. Helle N, Barkmann C, Bartz-Seel J, et al. Very low birth-weight as a risk factor for postpartum depression four to six weeks postbirth in mothers and fathers: Cross-sectional results from a controlled multicentre cohort study. *J Affect Disord*. 2015;180:154-161.

26. Howard LM, Ryan EG, Trevillion K, et al. Accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy. *Br J Psychiatry*. 2018;212(1):50-56.

27. Leonardou AA, Zervas YM, Papageorgiou CC, et al. Validation of the Edinburgh Postnatal Depression Scale and prevalence of postnatal depression at two months postpartum in a sample of Greek mothers. *J Reprod Infant Psychol*. 2009;27(1):28-39.

28. Nakić Radoš S, Tadinac M, Herman R. Validation study of the Croatian version of the Edinburgh Postnatal Depression Scale (EPDS). *Suvrem Psihol*. 2013;16(2):203-218.

29. Phillips J, Charles M, Sharpe L, Matthey S. Validation of the subscales of the Edinburgh Postnatal Depression Scale in a sample of women with unsettled infants. *J Affect Disord.* 2009;118(1-3):101-112.

30. Prenoveau J, Craske M, Counsell N, et al. Postpartum GAD is a risk factor for postpartum MDD: the course and longitudinal relationships of postpartum GAD and MDD. *Depress Anxiety*. 2013;30(6):506-514.

31. Quispel C, Schneider TA, Hoogendijk WJ, Bonsel GJ, Lambregtse-van den Berg MP. Successful five-item triage for the broad spectrum of mental disorders in pregnancy - a validation study. *BMC Pregnancy Childbirth.* 2015;15(1):51.

32. Rochat TJ, Tomlinson M, Newell ML, Stein A. Detection of antenatal depression in rural HIV-affected populations with short and ultrashort versions of the Edinburgh Postnatal Depression Scale (EPDS). *Arch Womens Ment Health*. 2013;16(5):401-410.

33. Stewart RC, Umar E, Tomenson B, Creed F. Validation of screening tools for antenatal depression in Malawi—A comparison of the Edinburgh Postnatal Depression Scale and Self Reporting Questionnaire. *J Affect Disord*. 2013;150(3):1041-1047.

34. Tandon SD, Cluxton-Keller F, Leis J, Le HN, Perry DF. A comparison of three screening tools to identify perinatal depression among low-income African American women. *J Affect Disord*. 2012;136(1-2):155-162.

35. Tendais I, Costa R, Conde A, Figueiredo B. Screening for depression and anxiety disorders from pregnancy to postpartum with the EPDS and STAI. *Span J Psychol*. 2014;17.

36. Töreki A, Andó B, Keresztúri A, et al. The Edinburgh Postnatal Depression Scale: translation and antepartum validation for a Hungarian sample. *Midwifery*. 2013;29(4):308-315.

37. Töreki A, Andó B, Dudas RB, et al. Validation of the Edinburgh Postnatal Depression Scale as a screening tool for postpartum depression in a clinical sample in Hungary. *Midwifery*. 2014;30(8):911-918.

38. Tran TD, Tran T, La B, Lee D, Rosenthal D, Fisher J. Screening for perinatal common mental disorders in women in the north of Vietnam: a comparison of three psychometric instruments. *J Affect Disord*. 2011;133(1-2):281-293.

39. Turner K, Piazzini A, Franza A, Marconi AM, Canger R, Canevini MP. Epilepsy and postpartum depression. *Epilepsia.* 2009;50:24-27.

40. Vega-Dienstmaier JM, Mazzotti GS, Campos MS. Validation of a Spanish version of the Edinburgh postnatal depression scale. *Actas Esp Psiquiatr*. 2002;30(2):106-111.

41. Statistics Canada. Canadian Community Health Survey - mental health and well-being. Ottawa (ON). 2011 [updated 2013 Sep 10; cited 2014 Sep 6]. Available from: http://www23.statcan.gc.ca:81/imdb/p2SV.pl?Function=getSurvey&SDDS=5015&lang=en&db=imdb&adm=8&dis=2.

42. Eaton WW, Anthony JC, Gallo J, et al. Natural history of Diagnostic Interview Schedule/DSM-IV major depression. The Baltimore Epidemiologic Catchment Area follow-up. *Arch Gen Psychiatry*. 1997;54(11):993-999.

43. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA.* 2003;289(23):3095-3105.

44. Weissman MM, Bland RC, Canino GJ, et al. Cross-national epidemiology of major depression and bipolar disorder. *JAMA.* 1996; 276(4):293-299.

45. Andrade L, Caraveo-Anduaga JJ, Berglund P, et al. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *Int J Methods Psychiatr Res*. 2003;12(1):3-21.

Figure 1. Flow diagram of the study selection process

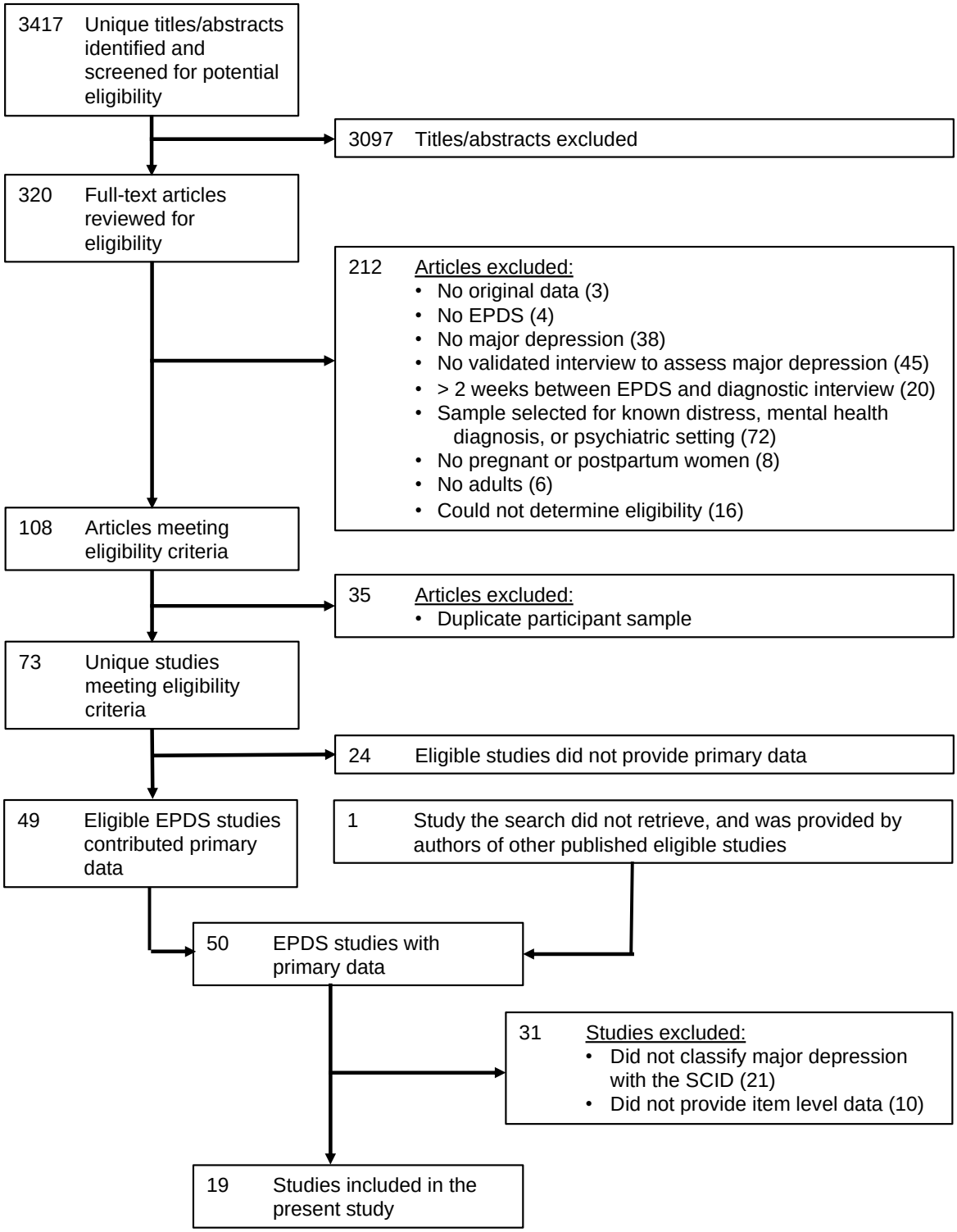

Table 1. Difference between EPDS-5 ≥ 7 prevalence and SCID prevalence for each included study

Author, Year	Country	N Total	N (%) EPDS-5 ≥ 7	N (%) SCID Major Depression	% Difference EPDS-5 ≥ 7 – SCID Major Depression	Ratio: EPDS-5 ≥ 7/ SCID Major Depression
Barnes, 2009[22]	UK	347	71 (20.5)	25 (7.2)	13.3	2.8
Beck, 2001[23]	USA	150	20 (13.3)	18 (12.0)	1.3	1.1
de Figueiredo, 2015[24a]	Brazil	242	94 (27.5)	95 (29.6)	-2.1	0.9
Helle, 2015[25]	Germany	225	42 (18.7)	12 (5.3)	13.3	3.5
Howard, 2018[26a]	UK	532	173 (17.0)	130 (9.4)	7.6	1.8
Leonardou, 2009[27]	Greece	81	13 (16.0)	4 (4.9)	11.1	3.3
Nakić Radoš, 2013[28]	Croatia	272	32 (11.8)	10 (3.7)	8.1	3.2
Phillips, 2009[29]	Australia	158	70 (44.3)	42 (26.6)	17.7	1.7
Prenoveau, 2013[30a]	UK	220	51 (14.7)	20 (6.0)	8.7	2.5
Quispel, 2015[31]	The Netherlands	31	0 (0.0)	0 (0.0)	0.0	-----
Rochat, 2013[32]	South Africa	104	66 (63.5)	50 (48.1)	15.4	1.3
Stewart, 2013[33a]	Malawi	186	46 (11.2)	34 (10.1)	1.1	1.1
Tandon, 2012[34]	USA	89	34 (38.2)	25 (28.1)	10.1	1.4
Tendais, 2014[35a]	Portugal	141	29 (10.9)	18 (7.6)	3.3	1.4
Töreki, 2013[36]	Hungary	219	6 (2.7)	7 (3.2)	-0.5	0.9
Töreki, 201[37]	Hungary	265	20 (7.5)	8 (3.0)	4.5	2.5
Tran, 2011[38]	Vietnam	361	28 (7.8)	53 (14.7)	-6.9	0.5
Turner, 2009[39]	Italy	29	2 (6.9)	2 (6.9)	0.0	1.0
Vega-Dienstmaier, 2002[40]	Peru	306	148 (48.4)	19 (6.2)	42.2	7.8
Pooled Results (with 95% confidence interval)		3,958	9.2% (6.0% to 13.7%)	16.2% (10.7% to 23.8%)	8.0% (2.9% to 13.2%)	Mean = 2.1 Median = 1.4

[a]Sampling weights were applied. Counts are based on actual numbers whereas percentages are weighted

Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; SCID: Structured Clinical Interview for DSM; UK: United Kingdom; USA: United States of America